Research Article

How can integrated morphotaxonomy- and metabarcoding-based diatom assemblage analyses best contribute to the ecological assessment of streams?

Panayiota Pissaridou[1], Marco Cantonati[2], Agnes Bouchez[3],
Iakovos Tziortzis[4], Gerald Dörflinger[4], Marlen I. Vasquez[1]

1 *Department of Chemical Engineering, Cyprus University of Technology, Archiepiskopou Kyprianou 30, Limassol, 3036, Cyprus*

2 *Museo delle Scienze – MUSE, Limnology and Phycology Section, Trento, Italy*

3 *INRAE, UMR CARRTEL, Thonon-les-bains, F-74200, France*

4 *Water Development Department, Kennedy Avenue 100-110, 1047 Pallouriotissa, Cyprus*

Corresponding author: Marlen I. Vasquez (marlen.vasquez@cut.ac.cy)

Abstract

Environmental conditions, such as nutrient concentrations, salinity, elevation etc., shape diatom assemblages of periphytic biofilms. These assemblages respond rapidly to environmental changes, a fact which makes diatoms valuable bioindicators. Hence, freshwater biomonitoring programmes currently use diatom indices (e.g. EU Water Framework Directive - WFD). To date, microscopy-based assessments require high taxonomic expertise for diatom identification at the species level. High-throughput technologies now provide cost-effective identification approaches that are promising, complementary or alternative tools for bioassessment. The suitability of the metabarcoding method is evaluated for the first time in the Cyprus streams WFD monitoring network, an eastern Mediterranean country with many endemic species and results are compared to the results acquired from the morphotaxonomic analysis. Morphotaxonomic identification was conducted microscopically, using the most updated taxonomic concepts, literature and online resources. At the same time, DNA metabarcoding involved the use of the *rbc*L 312 bp barcode, high-throughput sequencing and bioinformatic analysis. The ecological status was calculated using the IPS Index. Results show a positive correlation between morpho-taxonomic and molecular IPS scores. Discrepancies between the two methodologies are related to the limitations of both techniques. This study confirmed that *Fistulifera saprophila* can have a crucial role in key differences observed, as it negatively influences IPS scores and microscopy methods frequently overlook it. Importantly, gaps in the DNA barcoding reference databases lead to a positive overestimation in IPS scores. Overall, we conclude that DNA metabarcoding offsets the morphotaxonomic methodology for the ecological quality assessment of freshwaters.

Key Words

DNA metabarcoding, morpho-taxonomy, river biomonitoring, Cyprus rivers

Introduction

Freshwater ecosystems are transient environments, featuring hotspots of biodiversity and supporting about 10% of the total known species on the earth (Strayer and Dudgeon 2010). Due to the extensive urbanisation and advancement in human activities in recent decades, these ecosystems are subjected to environmental degradation and impacted by anthropogenic pressures (Hupało et al. 2021). As a result, these freshwater biodiversity hotspots are declining. The Mediterranean region fosters a unique combination of geography, geological history and climate, making it an excellent example of a freshwater biodiversity hotspot. This is estimated to result from a distinctly cool and wet season, followed by a warm and dry season influenced by a sequence of regular and often

PENSOFT.

extreme flooding and drying periods and relatively rare snowfall. Additionally, it is characterised by intense human activities. Mediterranean islands are no exception, being the most threatened ecosystems within the region, having intermittent and harsh-intermittent rivers highly impacted by human activities (Hupało et al. 2021). Although the Mediterranean is one of the most studied and well-known regions globally, little is known about the organisms inhabiting freshwater ecosystems. Studies focus mainly on vertebrates, macroinvertebrate orders and macrophytes of the northern and north-western domain (Figueroa et al. 2012).

Diatoms are unicellular photosynthetic autotrophs (microalgae) inhabiting a huge variety of habitats, such as oceans, rivers, wetlands and even the soil environment. They are the chief primary producers and key contributors in food webs, biogeochemical cycles and carbon fixation (Pérez-Burillo et al. 2020). Interestingly, an estimated 100,000 species of diatoms are believed to exist, with newly-discovered diatom species present in unique areas worldwide (Mann and Vanormelingen 2013). Diatom assemblages have selective growth preferences. They bloom at specific environmental conditions (salinity, temperature etc.) with short-generation periods (Smol and Stoermer 2010). Therefore, they respond rapidly to environmental changes that can influence their habitat, such as nutrient concentration, organic matter and water contamination (Hering et al. 2006).

For several decades, scientists have assessed the effect of pollution on diatom assemblages. Sixty years ago, scientists started assessing the influence of pollution on diatom assemblages and provided occurrence probabilities of microalgal species in water quality classes. During the late 20th century, scientists managed to incorporate the long-term monitoring of the diatom assemblages as a means of bioassessment (Butcher 1947; Zelinka and Marvan 1961; McCormick and Cairns 1994). In 2000, the Water Framework Directive (WFD) incorporated the concept of reference conditions and the adoption of phytobenthos as one of the obligatory Biological Quality Elements (BQEs) (European Comission 2000) considering that changes in diatom assemblages can be a valuable monitoring and assessment tool. Therefore, diatoms are used by the Member States to meet their legislative obligations towards the WFD (European Comission 2000).

To date, the quality of water-bodies worldwide is assessed by applying different diatom indices (Dalu et al. 2016). Such indices include the Biological Diatom Index (BDI), the Eutrophication Pollution Index (EPI), the Trophic Diatom Index (TDI), the Specific Polluosensitivity Index (IPS), the Sladecek Index (SLA) and the Saprobic Index of Rott (ROT), each of which has different sensitivities and investigate various environmental drives towards pollution (Rimet et al. 2005; Dalu et al. 2016; Pawlowski et al. 2018).

In Cyprus, the IPS Index (CEMAGREF 1982) is used since it was more suitable for the pressures that could affect the water quality in Cyprus' streams (WDD 2008)

and the official assessment method's class boundary values, based on the morphotaxonomy. The IPS Index correlates parameters related to organic pollution, ionic strength and eutrophication and, therefore, provides an overall estimate of water quality and is used by several European countries, such as Spain, Belgium, Estonia, Luxembourg, Sweden, Bulgaria, Greece and Portugal (WDD 2014; European Comission 2018; Masouras et al. 2021). It is worth mentioning that most of these indices and assessment concepts were developed elsewhere in Europe. Therefore their transfer to Cyprus and the eastern Mediterranean countries should be done with caution due to the combination of harsh climate and geographical isolation (Cantonati et al. 2020). In 2014, several Mediterranean countries performed an intercallibration exercise to provide a common framework for the successful comparison of diatom assessments of river quality across the Mediterranean European region (Almeida et al. 2014).

Diatom taxonomic classification involves morphological identification under the microscope by categorising the siliceous frustules of diatoms (Pandey et al. 2017). This is a time-consuming procedure that requires an accurate morphological identification at species or sub-species level. Therefore, it requires high taxonomic expertise (Kahlert et al. 2009). At the same time, such practices are subjected to biases and are susceptible to human error. New strategies have been examined to develop high-throughput, cost-effective approaches (Hering et al. 2018; Kelly et al. 2018; Vasselon et al. 2018). Recently, DNA metabarcoding, combined with high throughput sequencing and bioinformatics, is an alternative approach to morpho-taxonomic identification (Hering et al. 2018; Pawlowski et al. 2018; Rivera et al. 2018; Vasselon et al. 2018; Tapolczai et al. 2019). DNA metabarcoding involves identifying the microbial diversity from environmental samples by analysing relatively conserved small gene sequences (i.e. barcodes).

The start of Cyprus streams' bioassessment using diatoms dates back to 2005 (Cantonati et al. 2020) and the Water Development Department (WDD) of Cyprus has incorporated it into routine assessments since 2010. The routine assessment involves the use of the morpho-taxonomic identification method for the calculation of IPS. To date, DNA metabarcoding has been assessed mainly by the west-northern Mediterranean and northern European countries (Vasselon et al. 2017b; Kelly et al. 2018; Bailet et al. 2019; Feio et al. 2020; Pérez-Burillo et al. 2020). Additionally, extensive studies have been performed throughout Mediterranean marine environments regarding diatom biodiversity using DNA metabarcoding (Malviya et al. 2016; Kaleli et al. 2020; Vijver et al. 2020). Therefore, this study compares, for the first time, the diatom diversity and IPS scores obtained using high-throughput DNA sequencing and the morpho-taxonomic identification approach for an eastern Mediterranean country, like Cyprus, with a known population of endemic diatom species (Cantonati et al. 2016, 2018). Here, the study uses the results of the metabarcoding analysis

that utilised a small 312 bp fragment of the *rbc*L, which encodes the Ribulose-1,5-bisphosphate carboxylase/oxygenase. This gene shows alternating highly conserved and polymorphic regions, which are key requirements for a successful genetic identification to the species level (Kermarrec et al. 2013). We compared both methods to: a) assess whether the two methodologies are complementary or they can replace one another and b) test which methodologies are more suitable within Cyprus' routine bioassessment protocol using the IPS Index.

Materials and methods

Study area and sampling methodology.

Sampling stations were selected following the annual schedule of the WDD during May and June 2018. Biofilm samples were collected from 32 river sites in 11 catchments. Permanently submerged cobbles making up an area of about 100 cm² were randomly selected from the main stream channel in moderate flow conditions and a minimum of 5 cm water depth. The cobbles were gently rubbed with a brush at the parts exposed to the river flow. The dislodged material was preserved in formaldehyde (3%) for morphological or ethanol (70%) for the molecular analysis (CEN 2018). Samples were stored in a cool, dark area until laboratory analysis, as previously described by Pissaridou et al. (2021).

DNA extraction and molecular analysis

DNA extraction was performed using a commercial kit (NucleoSpin Soil Kit, MACHEREY-NAGEL) according to the manufacturer's instructions. Two technical samples were evaluated. Briefly, 2 ml of suspended diatom biofilm were centrifuged for 30 min at 13,000 rpm. The pellet was then lysed and the DNA was purified following the Kit's protocol. The DNA quality was assessed spectrophotometrically (NanoDrop 1000, Thermo Fisher Scientific). The *rbcL* barcode was amplified by PCR using diatom-specific *rbcL* primers and the HiFi KAPA Taq Mix (KAPA Biosciences) to a final volume of 25 ml. Samples were then sent to GenoToul Genomics and Transcriptomics platform (GeT-PlaGe, Auzeville, France) for Illumina MiSeq sequencing.

Bioinformatic analysis

Raw Fastq sequences were analysed using Mothur software (version 1.41.3) (Schloss et al. 2009). Sequences were filtered for quality and length (criteria: minimum length 280 bp, one mismatch allowed and homopolymers less than 8 bp), dereplicated and aligned to the Diat. barcode database version 7.1 (Rimet et al. 2019, 2018). Then, chimera detection and removal were performed. Taxonomic affiliation of reads was achieved using the Wang Method at a 75% confidence threshold. Lastly, a

distance matrix was created to allow the clustering of reads into operational taxonomical units (OTUs). The clustering method was the Furthest-Neighbour, where sequences with more than or equal to 95% similarity were clustered in one OTU. OTUs singletons were removed. Then all samples were normalised to the lowest read abundance. The method is described in detail in Pissaridou et al. (2021).

Morpho-taxonomic analysis

The diatom biofilm samples underwent processing using the hydrogen peroxide and potassium dichromate method, following the European standard EN 13946 (DIN EN 13946 2014), before permanent slide mounting. Briefly, 5 to 10 ml of diatom samples were centrifuged. Samples were then cleaned by repeated rinsing with 30% hydrogen peroxide (H_2O_2) and hydrochloric acid (HCl) following final decantation with distilled water. Next, samples were air-dried and mounted on microscope glass slides using Naphrax (Brunel Microscopes Ltd). At least 400 valves were identified for each sample and counted under an optical microscope Zeiss Axioskop 2, equipped with phase contrast and a digital camera Axiocam (Carl Zeiss JSC, Milan, Italy). The species identification was performed to the lowest taxonomic level possible (species or intraspecific rank) using various keys, with Cantonati et al. (2017) as the primary reference and recent scientific publications for specific taxa. The results were harmonised according to the "Final harmonisation of diatom names" list (Feio 2011). A complete species list and number of valves per species were created for each sample, along with each species' relative abundance.

Data analysis

Correction factor: The Correction Factor (CF) was applied to account for the *rbc*L copy number in relation to the biovolume of cells. The application was performed as described by Vasselon et al. (2018) according to the information supplied in the Diat.barcode database version 7.1 (Rimet et al. 2018, 2019).

***IPS score calculation*:**

$$IPS = \frac{\left(\sum_{i=1}^{n} \alpha_i \times IPSV_i \times IPSS_i\right) + \left(\sum_{i=1}^{n} \alpha_j \times IPSV_j \times IPSS_j\right) + \cdots}{\left(\sum_{i=1}^{n} \alpha_i \times IPSV_i\right) + \left(\sum_{i=1}^{n} \alpha_j \times IPSV_j\right) + \cdots}$$

IPSV = pollution tolerance values
IPSS = pollution sensitivity values

The IPSV and the IPSS values for each species were acquired from the Diat.barcode database version 7.1 (Rimet et al. 2018, 2019). IPS scores were transformed to water quality estimates where IPS ≥ 17 – "High"; IPS (13–17) – "Good"; IPS (9–13) – "Moderate"; IPS (5–9) – "Poor"; and IPS (1–5) – "Bad" class, respectively (CEMAGREF 1982).
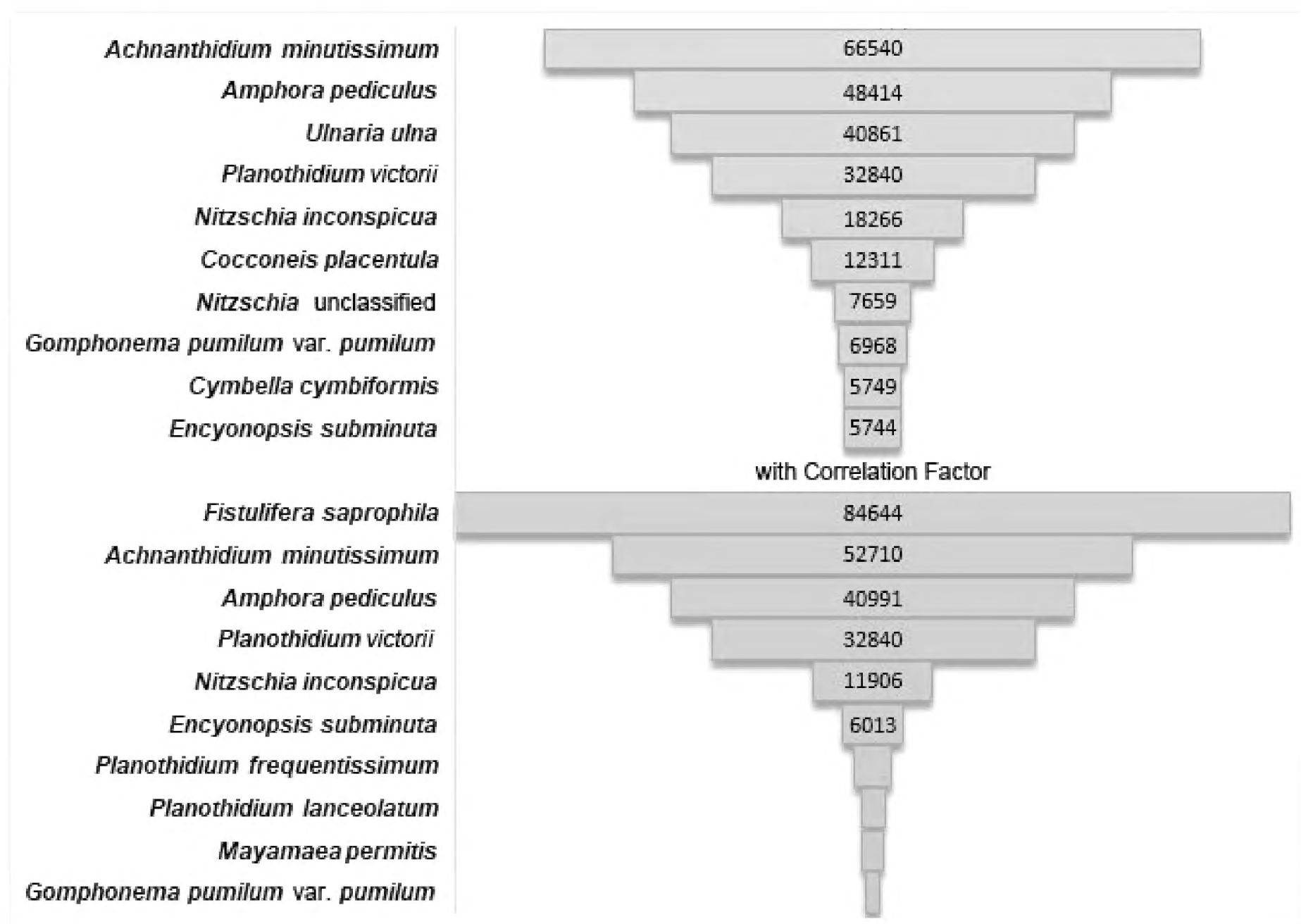
Figure 1. Correction factor application. Ten most abundant species identified by the molecular approach before and after the CF application.

Statistical analysis: Pearson's Correlation was performed using GraphPad PRISM version 9 for Windows, GraphPad Software, (La Jolla, California, USA). Pie charts and Venn Diagrams were performed in Excel (Microsoft). Plots were created using R (R Core Team 2020).

Results

DNA metabarcoding diatom identification

The DNA metabarcoding analysis identified a total of four classes, 13 orders, 28 families, 53 genera, 106 species and 136 strains amongst 1,553,538 reads for the 32 stations studied. Vasselon et al. (2018) observed that the molecular approach results overestimate the presence of high biovolume species, which may have higher *rbc*L copies per cell. The metabarcoding approach performed in this study was also able to observe such changes in abundance, as expressed as OTU copy numbers (Fig. 1). Consequently, a Correction Factor (CF) was applied to take into account the biovolume of diatom frustules to make DNA metabarcoding counts fit better with those estimated by microscopic counts (Charles et al. 2021). This removed species with high biovolume from the top ten most abundant species, like *Ulnaria ulna, Cocconeis euglypta* and *Cymbella cymbiformis* and, instead, intro-

duced species like *Fistulifera saprophila, Planothidium frequentissimum, Planothidium lanceolatum* and *Mayamaea permitis* with low biovolume (Fig. 1). This correction allows for a better alignment of the diatom assemblages from the molecular approach and subsequently a higher correlation between the two methodologies.

The ten most abundant species of the morpho-taxonomic analysis are represented in Figure 2. *Cocconeis euglypta, Achnanthidium jackii* and *A. minutissimum* are the most abundant diatoms. This was also the case in the molecular analysis (Fig. 1), as *A. minutissimum* was detected second in line. *A. jackii* was not detected in the molecular data as, in the reference database, it is considered a variant of *A. minutissimum* and is, therefore, represented by the same barcode (Rimet et al. 2019).

Comparison of the morpho-taxonomic and molecular identification of diatoms

To evaluate the performance of the two methods, we compared the species identified in both cases. Figure 3A shows that both methodologies identified 39 common species, corresponding to 18% (Suppl. material 1: Table S1). These 39 species represent a large part of the community in terms of biomass since the 39 species include several of the most frequent and abundant taxa in the Cyprus diatom assemblages. Interestingly, 99 species
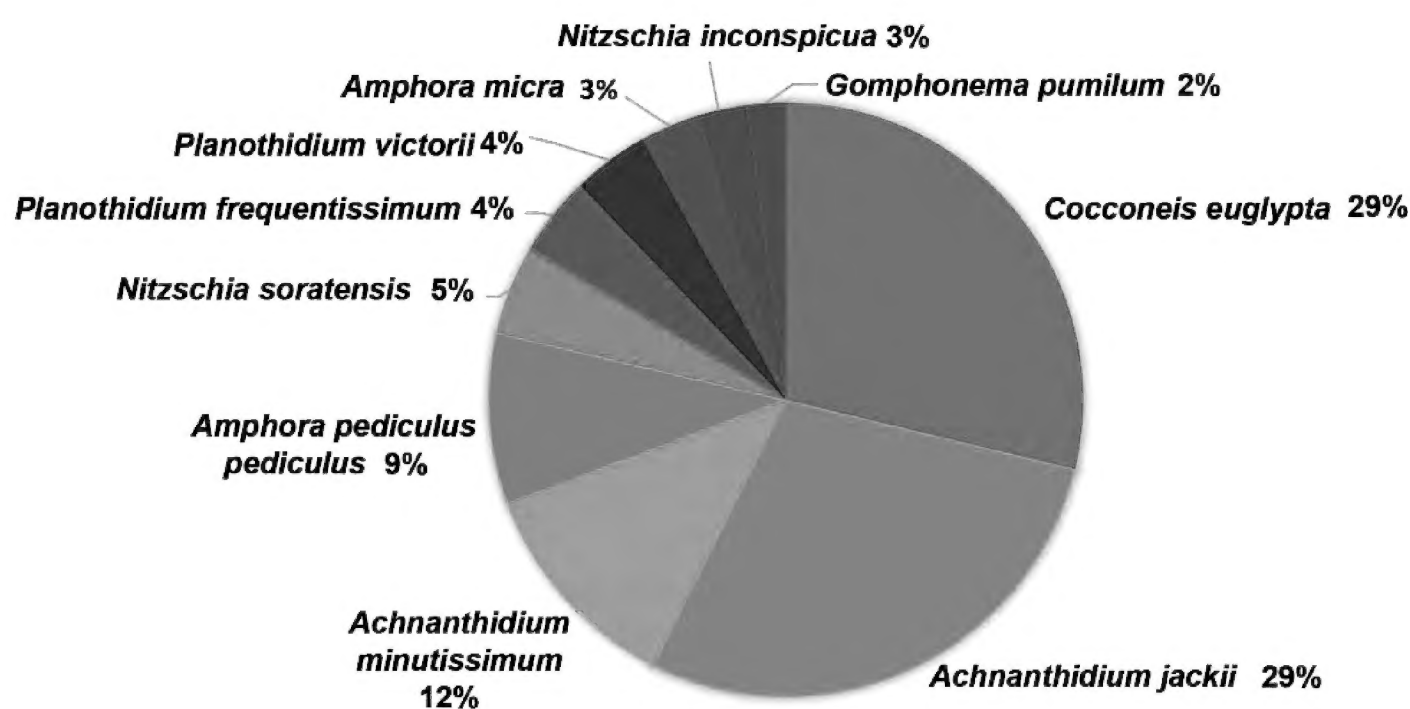
Figure 2. Ten most abundant species identified by the morpho-taxonomic approach.
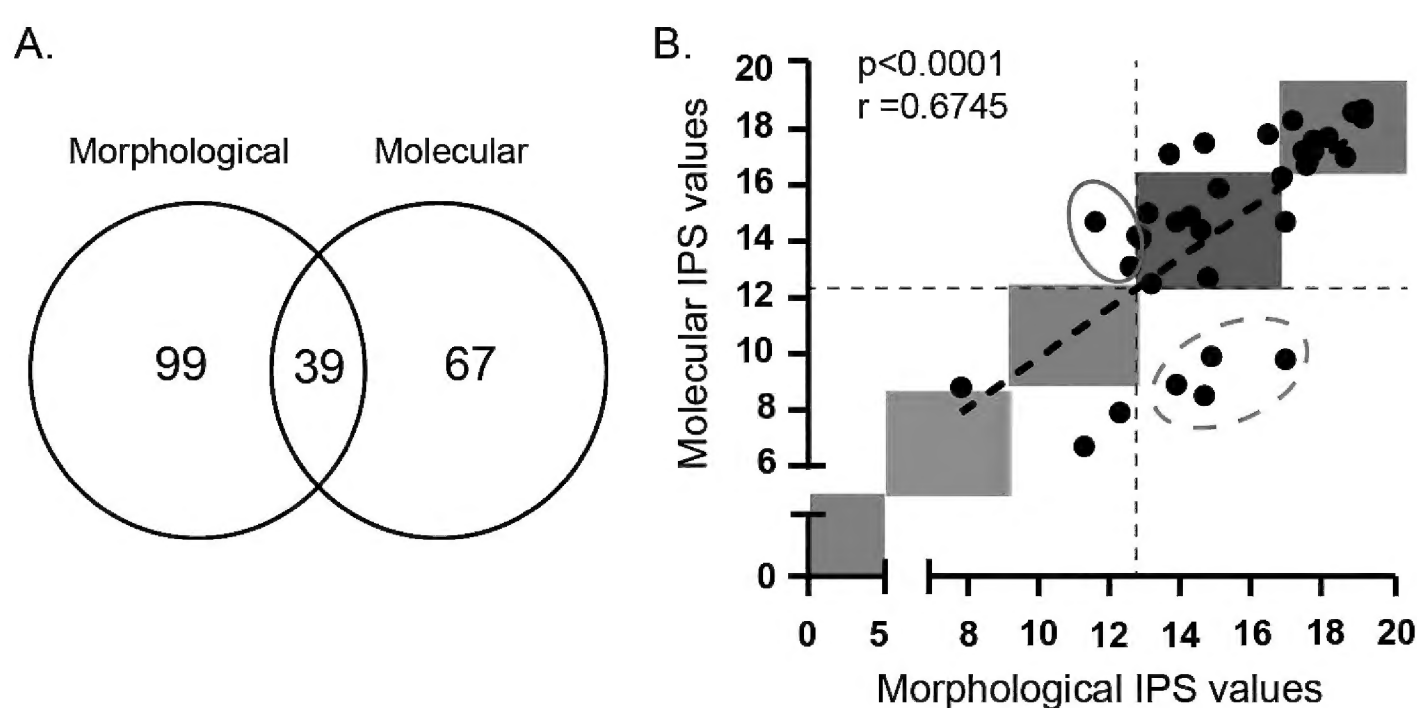
Figure 3. Comparison of morpho-taxonomic and molecular identification at the species level. **A)** Venn Diagram demonstrating the number of species identified by both methodologies, by morpho-taxonomic identification only (99), by molecular identification only (67) and by both methods (39). Suppl. material 1: Table S1 gives the number of valves and number of reads recorded for the 39 common species; **B)** IPS correlation of the morpho-taxonomic and molecular identification methodologies. Pearson Correlation was performed indicating a significant correlation ($r = 0.6745$, $p < 0.0001$. The boundaries of the ecological status classes are marked: red (Bad), orange (Poor), yellow (Moderate), green (Good), blue (High). Dashed lines: Good/moderate boundary, Red dashed circle: morphological classifies as good status, while molecular classifies below good status. Blue cirlce: morphological classifies as below good status, while molecular classifies good status.

were identified only by the morpho-taxonomic identification and 67 species only by the molecular identification method (Fig. 3A).

The number of species observed with each technique was used to calculate the IPS index for the ecological status evaluation of each station (CEMAGREF 1982). The IPS scores, obtained with the two methodologies, show significant positive correlation ($p < 0.0001$, $r = 0.6745$, Fig. 3B). More than half of the samples shared the same quality class.

At the same time, some discrepancies between the two methodologies were observed. Twelve stations show substantial differences between the two approaches (Fig. 3B, Fig. 4, Suppl. material 2). The molecular approach calculated lower ecological status values than the morpho-taxonomic approach for six stations (Fig. 3B). Interestingly, four out of six stations have *Fistulifera saprophila* as

the most abundant species. As detected by the molecular analysis, one station has *Amphora pediculus* as the most abundant with *Fistulifera saprophila* following and one station has *Achnanthidium minutissimum* and *Ulnaria ulna* as the most abundant species.

On the other hand, the molecular approach calculated higher IPS values than the morpho-taxonomic approach for six more stations (Suppl. material 2). These differences could be due to the lack of barcoding data for diatom species at these stations, creating gaps in comparing the two methodologies. The molecular approach identifies a higher abundance of *Planothidium victorii*, *Hantzschia* unclassified and *Achnanthes* unclassified. For the same stations, higher abundance of *Achnanthidium jackii*, *Achnanthidium minutissimum*, *Cocconeis euglypta*, *Cyclostephanos* cf. *invisitatus*, *Amphora pediculus* and *Amphora micra* have been identified
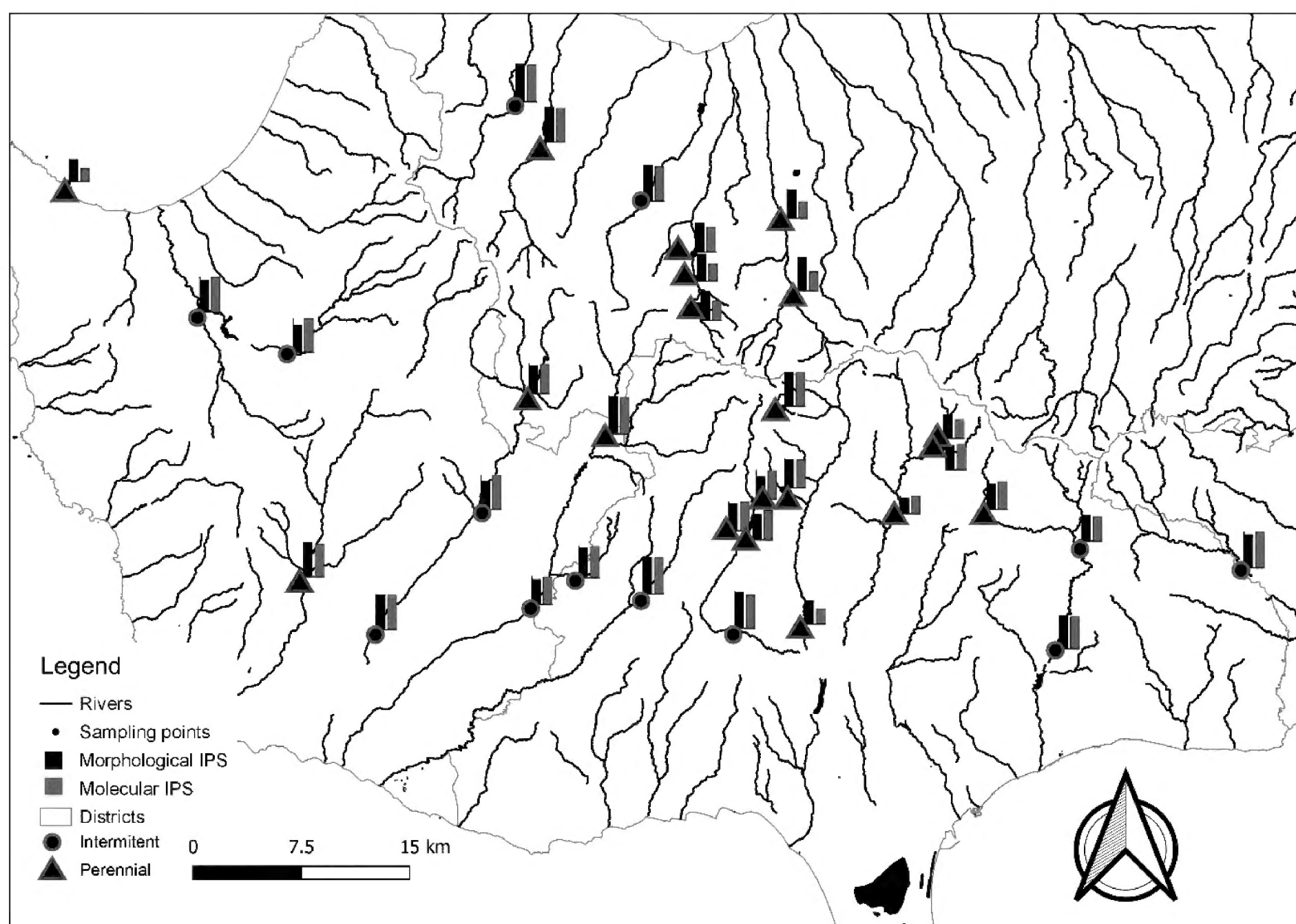
Figure 4. Map representation and comparison of the IPS values obtained using the morpho-taxonomic and molecular approach. The detailed values are provided in Suppl. material 2: Fig. S1.

using the morpho-taxonomic approach (data not shown) (Pérez-Burillo et al. 2020). As previously noted, *Cyclostephanos* cf. *invisitatus* and *Amphora micra* are diatoms species whose barcode is not yet available in online barcode databases.

Discussion

Thus far, several studies have shown disagreements between molecular and morphological datasets due to the inherent biases of both methods (Keck et al. 2021). The most apparent reason for these discrepancies is the incompleteness of molecular reference databases since barcoding studies focused so far on particular river types and geographical regions, including mostly north-west Mediterranean countries and perennial rivers. (Pawlowski et al. 2018).

Here, as presented by recent studies, the identification of more species by the morphological approach than by DNA metabarcoding, in many cases, does not significantly impact the IPS assignment (Vasselon et al. 2017a; Keck et al. 2018; Rivera et al. 2018; Mortágua et al. 2019). Additionally, we have demonstrated how the CF's application to the molecular approach is important in comparing the two methodologies. The CF accounts for the *rbc*L copy number concerning the biovolume of cells

and, therefore, is considered to obtain DNA sequence proportions closer to cell proportions obtained through microscopy and account for large diatom frustules having several *rbc*L copies (Vasselon et al. 2019). Both methodologies have their limitations, including the classification biases in species complexes and oversimplification requiring high scientific expertise for morphology-based approaches and the gaps of DNA barcoding data in existing references databases in molecular-based approaches (Jahn et al. 2017; Pinseel et al. 2017). Kelly et al. (2020) suggested that, as microscopy data also have some biases and limitations, it would be better to treat molecular data at their face value to integrate them into indicator metrics. Indeed, the differences in size reflected by the *rbc*L data (Vasselon et al. 2018) may give a more accurate indication of the relative contribution of each taxon to diatom productivity than the microscopy data, which could lead to novel bioindicator metrics, better reflecting the ecosystem functioning.

In this paper, some discrepancies in IPS scores appear when comparing both approaches. The microscopy can sometimes underestimate small diatom species with low sensitivity values, thereby having a significant negative impact on the IPS scores estimated with DNA metabarcoding (Jahn et al. 2017; Pinseel et al. 2017). Interestingly, literature suggests that small species like *Fistulifera saprophila* and *Mayamaea permitis* are often overlooked/

underestimated by the morphological approach (Kahlert et al. 2009; Pérez-Burillo et al. 2020). In this study, only *F. saprophila* was detected exclusively by the molecular approach, whereas *M. permitis* was detected by both methods (Suppl. material 1). Both species become prominent in eutrophic and polluted environments and consequently lower the IPS Index scores (Rott et al. 1998; Rimet et al. 2005; Kahlert et al. 2009; Pérez-Burillo et al. 2020). This explains why six of the twelve differing stations have lower IPS scores when estimated through the molecular approach (Suppl. material 2). *F. saprophila* valves are known to be prone to dissolution during laboratory preparation and this could explain why this species is usually overlooked in the morphological identification procedures (Zgrundo et al. 2013; Cantonati et al. 2017, 2018; Kelly et al. 2020). Our results agree with the observations of Pérez-Burillo et al. (2020) that *Fistulifera saprophila* was often the species responsible for sites becoming critical, i.e. having a low IPS score. The same study identified *A. minutissimum* as the diatom species contributing positively to the IPS scores (Pérez-Burillo et al. 2020).

On the other hand, gaps in the reference databases may be unable to identify some other taxa of low sensitivity value, with a significant positive impact on the IPS scores estimated with DNA metabarcoding (Vasselon et al. 2019; Bailet et al. 2020). Regarding the eastern Mediterranean countries, there is a lack of sufficient DNA barcoding studies to date. Most recent DNA barcoding and metabarcoding studies focus on the west-northern Mediterranean and northern European countries (Vasselon et al. 2017b; Kelly et al. 2018; Bailet et al. 2019; Feio et al. 2020; Pérez-Burillo et al. 2020). When considering the eastern Mediterranean region, most studies focus on the marine environment (Malviya et al. 2016; Kaleli et al. 2020; Vijver et al. 2020) and morphological identification of endemic species, such as *Navicula veronensis* sp. nov. found in low elevation carbonate springs (Cantonati et al. 2016), *Ulnaria monodii* found in streams with relatively warm waters and *Ulnaria acuscypriacus* sp. nov. having tolerance to brackish conditions (Cantonati et al. 2018). Therefore, the gaps, observed here in the comparison of the two methodologies, are more or less expected.

Even though the morpho-taxonomic approach is able to identify more species than the DNA metabarcoding method, the DNA metabarcoding approach is still considered an effective method for diatom assemblage identification as it offers high-throughput, cost-effective and accurate analysis provided that the online barcode databases are up to date (Vasselon et al. 2017a; Keck et al. 2018; Rivera et al. 2018; Mortágua et al. 2019).

Conclusions

To conclude, this study has investigated diatom samples from the Cyprus stream monitoring network and the correlation of the two methodologies currently in play for stream-quality bioassessments, based on phytobenthos/diatoms. For the first time, we compared and contrasted DNA metabarcoding to the traditional morphological, ecological assessment for river biomonitoring in Cyprus, a typical eastern Mediterranean country. Our results are in line with what has already been shown in literature. The limitations of both techniques can offset each other nicely when using spatial datasets. Despite the differences between the two methods and the taxonomic inventories they produced, the IPS scores from the molecular data were well correlated with those from morpho-taxonomic data when a CF is applied to bring molecular data in line with the expectations, based on microscopy data. As Kelly et al. (2020) stated: "HTS is not a simple transaction in which LM is replaced by a better technology, but requires a deeper level of institutional transformation than had hitherto been anticipated".

Future work can investigate how the two methodologies could contribute to an integrated assessment of different Mediterranean rivers with different hydrological impacts and pollution (Charles et al. 2021). Additionally, it is worth studying the temporal observation of rivers to examine whether a potential shift to metabarcoding alters perceptions of status at an individual site level, as previously demonstrated by Kelly et al. (2020). Lastly, there is a parallel need of: 1) an enrichment of DNA reference databases for the diatom species that are primarily present and endemic to southern-eastern Mediterranean rivers and 2) development of so-called taxonomy-free approaches to overcome these reference gaps, as recently tested on diatoms (Apothéloz-Perret-Gentil et al. 2020; Tapolczai et al. 2021).

Acknowledgements

References

Almeida SFP, Elias C, Ferreira J, Tornés E, Puccinelli C, Delmas F, Dörflinger G, Urbanič G, Marcheggiani S, Rosebery J, Mancini L, Sabater S (2014) Water quality assessment of rivers using diatom metrics across Mediterranean Europe: A methods intercalibration exercise. Science of the Total Environment 476–477: 768–776. https://doi.org/10.1016/j.scitotenv.2013.11.144

Apothéloz-Perret-Gentil L, Bouchez A, Cordier T, Cordonier A, Guéguen J, Rimet F, Vasselon V, Pawlowski J (2020) Monitoring the ecological status of rivers with diatom eDNA metabarcoding: A comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. Molecular Ecology 30(13): 2959–2968. https://doi.org/10.1111/mec.15646

Bailet B, Apothéloz-Perret-Gentil L, Baričević A, Chonova T, Franc A, Frigerio JM, Kelly M, Mora D, Pfannkuchen M, Proft S, Ramon M, Vasselon V, Zimmermann J, Kahlert M (2020) Diatom DNA metabarcoding for ecological assessment: Comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. Science of the Total Environment 745: 140948. https://doi.org/10.1016/j.scitotenv.2020.140948

Bailet B, Bouchez A, Franc A, Frigerio J-M, Keck F, Karjalainen S-M, Rimet F, Schneider S, Kahlert M (2019) Molecular versus morphological data for benthic diatoms biomonitoring in Northern Europe freshwater and consequences for ecological status. Metabarcoding and Metagenomics 3: e34002. https://doi.org/10.3897/mbmg.3.34002

Butcher RW (1947) Studies in the ecology of rivers. IV. The algae of organically enriched water. Journal of Ecology 35: 186–191. https://doi.org/10.2307/2256507

Cantonati M, Kelly MG, Lange-Bertalot H (2017) Freshwater Benthic Diatoms of Central Europe. Academy of Natural Sciences of Drexel University: 1–901. http://nus.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwY2AwNtIz0EUrE4C1qKmxSTKw7QCaZjNLBHZ-CLIyBbZXUlETzpBTwbmknJ6PgKGNXb-MoxHpZYEsLaGg-p5ETB4nT4wIWxBajhYc7MwGQMuq7BMtwQtG3D0gRY-nYKawJCDdeB80HEiQEOQ6g03QQYW0F4CIQam1DxhBjanfG-BLrFKEQccN2MfNKAe28ooUkoCFfkZ [March 19, 2020]

Cantonati M, Angeli N, Spitale D, Lange-Bertalot H (2016) A new Navicula (Bacillariophyta) species from low-elevation carbonate springs affected by anthropogenic disturbance. Fottea 16: 255–265. https://doi.org/10.5507/fot.2016.013

Cantonati M, Lange-Bertalot H, Kelly MG, Angeli N (2018) Taxonomic and ecological characterization of two ulnaria species (Bacillariophyta) from streams in Cyprus. Phytotaxa 346: 78–92. https://doi.org/10.11646/phytotaxa.346.1.4

Cantonati M, Kelly MG, Demartini D, Angeli N, Dörflinger G, Papatheodoulou A, Armanini DG (2020) Overwhelming role of hydrology-related variables and river types in driving diatom species distribution and community assemblage in streams in Cyprus. Ecological Indicators 117: 106690. https://doi.org/10.1016/j.ecolind.2020.106690

CEMAGREF (1982) Etude des méthodes biologiques d'appréciation quantitative de la qualité des eaux. Bassin Rhone-Mediterranee-Corse: 1–218.

CEN (2018) CEN/TR 17245: Water quality – Technical report for the routine sampling of benthic diatoms from rivers and lakes adapted for metabarcoding analyses. CEN/TC 230/WG 23 – Aquatic Macrophytes and Algae: 1–8.

Charles DF, Kelly MG, Stevenson RJ, Poikane S, Theroux S, Zgrundo A, Cantonati M (2021) Benthic algae assessments in the EU and the US: Striving for consistency in the face of great ecological diversity. Ecological Indicators 121: 107082. https://doi.org/10.1016/j.ecolind.2020.107082

Dalu T, Bere T, Froneman PW (2016) Assessment of water quality based on diatom indices in a small temperate river system, Kowie River, South Africa. Water SA 42: 183–193. https://doi.org/10.4314/wsa.v42i2.02

DIN EN 13946 (2014) Water quality—guidance standard for routine sampling and pre-treatment of benthic diatoms from rivers.

European Comission (2000) Directive of the European Parliament and of the Council 2000/60/EC. Establishing a Framework for Community Action in the Field of Water Policy. Official Journal of the European Parliament L327: 1–82. https://doi.org/10.1039/ap9842100196

European Comission (2018) 2001 Euratom Establishing, pursuant to Directive 2000/60/EC of the European Parliament and of the Council, the values of the Member State monitoring system classifications as a result of the intercalibration exercise and repealing Commission Decision 2013/480/EU, 20–30.

Feio MJ (2011) WFD Intercalibration Phase 2: Milestone 6 report – River/Med GIG /Benthic Invertebrates (30th December 2011).

Feio MJ, Serra SRQ, Mortágua A, Bouchez A, Rimet F, Vasselon V, Almeida SFP (2020) A taxonomy-free approach based on machine learning to assess the quality of rivers with diatoms. Science of the Total Environment 722: 137900. https://doi.org/10.1016/j.scitotenv.2020.137900

Figueroa JMT de, López-Rodríguez MJ, Fenoglio S, Sánchez-Castillo P, Fochetti R (2012) Freshwater biodiversity in the rivers of the Mediterranean Basin. Hydrobiologia 719: 137–186. https://doi.org/10.1007/s10750-012-1281-z

Hering D, Borja A, Jones JI, Pont D, Boets P, Bouchez A, Bruce K, Drakare S, Hänfling B, Kahlert M, Leese F, Meissner K, Mergen P, Reyjol Y, Segurado P, Vogler A, Kelly M (2018) Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. Water Research 138: 192–205. https://doi.org/10.1016/j.watres.2018.03.003

Hering D, Johnoson KR, Kramm S, Schumtz S, Szoszkiewicz K, Verdonschot MPF (2006) Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: a comparative metric-based analysis of organism response to stress. Freshwater Biology 51: 1757–1785. https://doi.org/10.1111/j.1365-2427.2006.01610.x

Hupało K, Schmidt S, Macher T, Weiss M, Leese F (2021) Fresh insights into Mediterranean biodiversity: Environmental DNA reveals spatio-temporal patterns of stream invertebrate communities. bioRxiv. https://doi.org/10.1101/2021.06.10.447858

Jahn R, Abarca N, Gemeinholzer B, Mora D, Skibbe O, Kulikovskiy M, Gusev E, Kusber W-HH, Zimmermann J (2017) Planothidium lanceolatum and Planothidium frequentissimum reinvestigated with molecular methods and morphology: four new species and the taxonomic importance of the sinus and cavum. Diatom Research 32: 75–107. https://doi.org/10.1080/0269249X.2017.1312548

Kahlert M, Albert R-L, Anttila E-L, Bengtsson R, Bigler C, Eskola T, Gälman V, Gottschalk S, Herlitz E, Jarlman A, Kasperoviciene J, Kokociński M, Luup H, Miettinen J, Paunksnyte I, Piirsoo K, Quintana I, Raunio J, Sandell B, Simola H, Sundberg I, Vilbaste

S, Weckström J (2009) Harmonization is more important than experience—results of the first Nordic–Baltic diatom intercalibration exercise 2007 (stream monitoring). Journal of Applied Phycology 21: 471–482. https://doi.org/10.1007/s10811-008-9394-5

Kaleli MA, Kociolek JP, Solak CN (2020) Taxonomy and distribution of diatoms on the Turkish Mediterranean Coast, Dalyan (Mugla). Mediterranean Marine Science 21: 201–215. https://doi.org/10.12681/mms.17293

Keck F, Blackman RC, Bossart R, Brantschen J, Couton M, Hürlemann S, Kirschner D, Locher N, Zhang H, Altermatt F (2021) Meta-analysis shows both congruence and complementarity of DNA metabarcoding to traditional methods for biological community assessment. bioRxiv: 2021.06.29.450286. https://doi.org/10.1101/2021.06.29.450286

Keck F, Vasselon V, Rimet F, Bouchez A, Kahlert M (2018) Boosting DNA metabarcoding for biomonitoring with phylogenetic estimation of operational taxonomic units' ecological profiles. Molecular Ecology Resources 18: 1299–1309. https://doi.org/10.1111/1755-0998.12919

Kelly M, Boonham N, Juggins S, Kille P, Mann DG, Pass D, Sapp M, Sato S, Glover R (2018) Environment Agency A DNA based diatom metabarcoding approach for Water Framework Directive classification of rivers. 157 pp.

Kelly MG, Juggins S, Mann DG, Sato S, Glover R, Boonham N, Sapp M, Lewis E, Hany U, Kille P, Jones T, Walsh K (2020) Development of a novel metric for evaluating diatom assemblages in rivers using DNA metabarcoding. Ecological Indicators 118: 106725. https://doi.org/10.1016/j.ecolind.2020.106725

Malviya S, Scalco E, Audic S, Vincent F, Veluchamy A, Poulain J, Wincker P, Iudicone D, Vargas C de, Bittner L, Zingone A, Bowler C (2016) Insights into global diatom distribution and diversity in the world's ocean. Proceedings of the National Academy of Sciences 113: E1516–E1525. https://doi.org/10.1073/pnas.1509523113

Mann DG, Vanormelingen P (2013) An inordinate fondness? The number, distributions, and origins of diatom species. Journal of Eukaryotic Microbiology 60: 414–420. https://doi.org/10.1111/jeu.12047

Masouras A, Karaouzas I, Dimitriou E, Tsirtsis G, Smeti E (2021) Benthic diatoms in river biomonitoring-present and future perspectives within the Water Framework Directive. Water (Switzerland) 13(4): e478. https://doi.org/10.3390/w13040478

McCormick PV, Cairns J (1994) Algae as indicators of environmental change. Journal of Applied Phycology 6: 509–526. https://doi.org/10.1007/BF02182405

Mortágua A, Vasselon V, Oliveira R, Elias C, Chardon C, Bouchez A, Rimet F, João Feio M, Almeida SFP (2019) Applicability of DNA metabarcoding approach in the bioassessment of Portuguese rivers using diatoms. Ecological Indicators 106: 105470. https://doi.org/10.1016/j.ecolind.2019.105470

Pandey LK, Bergey EA, Lyu J, Park J, Choi S, Lee H, Depuydt S, Oh Y-T, Lee S-M, Han T (2017) The use of diatoms in ecotoxicology and bioassessment: Insights, advances and challenges. Water Research 118: 39–58. https://doi.org/10.1016/j.watres.2017.01.062

Pawlowski J, Kelly-Quinn M, Altermatt F, Apothéloz-Perret-Gentil L, Beja P, Boggero A, Borja A, Bouchez A, Cordier T, Domaizon I, Feio MJ, Filipe AF, Fornaroli R, Graf W, Herder J, van der Hoorn B, Iwan Jones J, Sagova-Mareckova M, Moritz C, Barquín J, Piggott JJ, Pinna M, Rimet F, Rinkevich B, Sousa-Santos C, Specchia V, Trobajo R, Vasselon V, Vitecek S, Zimmerman J, Weigand A, Leese

F, Kahlert M (2018) The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Science of The Total Environment 637–638: 1295–1310. https://doi.org/10.1016/j.scitotenv.2018.05.002

Pérez-Burillo J, Trobajo R, Vasselon V, Rimet F, Bouchez A, Mann DG (2020) Evaluation and sensitivity analysis of diatom DNA metabarcoding for WFD bioassessment of Mediterranean rivers. Science of The Total Environment 727: 138445. https://doi.org/10.1016/j.scitotenv.2020.138445

Pinseel E, Vanormelingen P, Hamilton PB, Vyverman W, Vijver B Van de, Kopalova K (2017) Molecular and morphological characterization of the Achnanthidium minutissimum complex (Bacillariophyta) in Petuniabukta (Spitsbergen, High Arctic) including the description of A. digitatum sp. nov. European Journal of Phycology 52: 264–280. https://doi.org/10.1080/09670262.2017.1283540

R Core Team (2020) R: A language and environment for statistical computing. https://www.r-project.org/

Rimet F, Gusev E, Kahlert M, Kelly MG, Kulikovskiy M, Maltsev Y, Mann DG, Pfannkuchen M, Trobajo R, Vasselon V, Zimmermann J, Bouchez A (2019) Diat.barcode, an open-access curated barcode library for diatoms. Scientific Reports 9: e15116. https://doi.org/10.1038/s41598-019-51500-6

Rimet F, Cauchie H-M, Hoffmann L, Ector L (2005) Response of diatom indices to simulated water quality improvements in a river. Journal of Applied Phycology 17: 119–128. https://doi.org/10.1007/s10811-005-4801-7

Rimet F, Chonova T, Gassiole G, Gusev E, Kahlert M, Keck F, Kelly M, Kochoska H, Kulikovskiy M, Levkov Z, Maltsev Y, Mann D, Pfannkuchen M, Trobajo R, Vasselon V, Wetzel C, Zimmermann J, Bouchez A (2018) Diat.barcode, an open-access barcode library for diatoms. Portail Data Inra. https://data.inra.fr/dataset.xhtml?persistentId=doi:10.15454/TOMBYZ [September 10, 2019]

Rivera SF, Vasselon V, Jacquet S, Bouchez A, Ariztegui D, Rimet F (2018) Metabarcoding of lake benthic diatoms: from structure assemblages to ecological assessment. Hydrobiologia 807: 37–51. https://doi.org/10.1007/s10750-017-3381-2

Rott E, Duthie HC, Pipp E (1998) Monitoring organic pollution and eutrophication in the Grand River, Ontario, by means of diatoms. Canadian Journal of Fisheries and Aquatic Sciences 55: 1443–1453. https://doi.org/10.1139/f98-038

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing Mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75: 7537–7541. https://doi.org/10.1128/AEM.01541-09

Smol JP, Stoermer EF [Eds] (2010) The Diatoms: Applications for the Environmental and Earth Sciences. Cambridge University Press, Cambridge, [xviii +] 667 pp. https://doi.org/10.1017/CBO9780511763175

Strayer DL, Dudgeon D (2010) Freshwater biodiversity conservation: Recent progress and future challenges. Journal of the North American Benthological Society 29: 344–358. https://doi.org/10.1899/08-171.1

Tapolczai K, Vasselon V, Bouchez A, Stenger-Kovács C, Padisák J, Rimet F (2019) The impact of OTU sequence similarity threshold on diatom-based bioassessment: A case study of the rivers of Mayotte

(France, Indian Ocean). Ecology and Evolution 9: 166–179. https://doi.org/10.1002/ece3.4701

Tapolczai K, Selmeczy GB, Szabó B, B-Béres V, Keck F, Bouchez A, Rimet F, Padisák J (2021) The potential of exact sequence variants (ESVs) to interpret and assess the impact of agricultural pressure on stream diatom assemblages revealed by DNA metabarcoding. Ecological Indicators 122: 107322. https://doi.org/10.1016/j.ecolind.2020.107322

Vasselon V, Rimet F, Tapolczai K, Bouchez A (2017a) Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte Island, France). Ecological Indicators 82: 1–12. https://doi.org/10.1016/j.ecolind.2017.06.024

Vasselon V, Domaizon I, Rimet F, Kahlert M, Bouchez A (2017b) Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? Freshwater Science 36: 162–177. https://doi.org/10.1086/690649

Vasselon V, Rimet F, Domaizon I, Monnier O, Reyjol Y, Bouchez A (2019) Assessing pollution of aquatic environments with diatoms' DNA metabarcoding: Experience and developments from France Water Framework Directive networks. Metabarcoding and Metagenomics 3: 101–115. https://doi.org/10.3897/mbmg.3.39646

Vasselon V, Bouchez A, Rimet F, Jacquet S, Trobajo R, Corniquel M, Tapolczai K, Domaizon I (2018) Avoiding quantification bias in metabarcoding: Application of a cell biovolume correction factor in diatom molecular biomonitoring. Mahon A (Ed.). Methods in Ecology and Evolution 9(4): 1060–1069. https://doi.org/10.1111/2041-210X.12960

Vijver B Van de, Robert K, Majewska R, Frankovich TA, Panagopoulou A, Bosak S (2020) Geographical variation in the diatom communities associated with loggerhead sea turtles (Caretta caretta). PLoS ONE 15: e0236513. https://doi.org/10.1371/journal.pone.0236513

WDD (2014) Report on the classification of water status (rivers, natural lakes, water reservoirs).

Zelinka M, Marvan P (1961) Zur Präzisierung der biologischen Klassifikation der Reinheit fliessender Gewässer. Archiv für Hydrobiologie 57: 389–407.

Zgrundo A, Lemke P, Pniewski F, Cox EJ, Latala A (2013) Morphological and molecular phylogenetic studies on Fistulifera saprophila. Diatom Research 28: 431–443. https://doi.org/10.1080/0269249X.2013.833136

Supplementary material 1
Table S1
Author: Panayiota Pissaridou, Marco Cantonati, Agnes Bouchez, Iakovos Tziortzis, Gerald Dörflinger, Marlen I Vasquez
Data type: docx. file
Explanation note: Diatom species recognised by both methodologies.
Link: https://doi.org/10.3897/mbmg.5.68438.suppl1

Supplementary material 2
Figure S1
Author: Panayiota Pissaridou, Marco Cantonati, Agnes Bouchez, Iakovos Tziortzis, Gerald Dörflinger, Marlen I Vasquez
Data type: docx. file
Explanation note: Relative abundance of diatoms species obtained by the molecular and the morpho-taxonomic approaches. Others correspond to diatoms species with less than 1% relative abundance.
Link: https://doi.org/10.3897/mbmg.5.68438.suppl2